DOCUMENT RESUME

ED 039 008 LI 001 935

AUTHOR Coulson, John
TITLE A Comparative Study of Differences in Judged Relevance of Abstracts Located by Two Information Systems.
INSTITUTION Research and Information Services for Education, King of Prussia, Pa.
PUB DATE [70]
NOTE 11p.

EDRS PRICE EDRS Price MF-$0.25 HC-$0.65
DESCRIPTORS Automation, Comparative Analysis, *Information Retrieval, *Information Systems, *Relevance (Information Retrieval), *Search Strategies, *Use Studies

ABSTRACT

Differences in judged relevance of Educational Resources Information Center (ERIC) abstracts identified by two different information systems were investigated. A national semi-private educational service agency which utilizes mechanical searching of the research literature was one information system, while the other was a regional, Federally-funded agency employing manual searching. Twenty clients reported they found substantially more ERIC abstracts identified by a manual method to be relevant to their needs than abstracts located by a mechanical searching system using the same set of ERIC descriptors. (Author)

ED039008

A COMPARATIVE STUDY OF DIFFERENCES IN JUDGED RELEVANCE OF ABSTRACTS LOCATED BY TWO INFORMATION SYSTEMS

John Coulson*

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Introduction

The number of information systems and centers in educational settings is increasing nationally and more are contemplated. Many of these information retrieval systems are automated and are intended to provide information to the user/educator quickly and efficiently. Whether or not these systems or centers are providing relevant information to requestors is not clear. To provide preliminary data which would begin to serve as a basis for rational decisions continues to be a problem in education.

To improve research information service, and to provide a foundation for valid judgments about the efficacy of providing relevant information via automated retrieval systems, a modest experimental pilot research study was undertaken. It was decided to conduct a comparative study of two different information retrieval systems using Educational Resources Information Center materials (E.R.I.C.). A national semi-private research information agency (N.I.A.)[1] which utilizes machine searching of the research literature was compared with the Research and Information Services for Education (R.I.S.E.), a regional research information agency employing manual searching of the research literature.

R.I.S.E., an agency funded under Title III of the Elementary and Secondary Education Act of 1965, is an educational research agency functioning at the local-intermediate-regional level. One of its more visible activities is to provide research information on request to educators. During the first two years of existence, over 800 requests for information were handled. In an effort to improve its information service, R.I.S.E. has continually reviewed and evaluated both the process and product of its efforts.

AC 001935

---

*John Coulson, Ed.D., is Associate Director, Research and Information Services for Education, 443 S. Gulph Rd., King of Prussia, Pa. 19406.

A short evaluation form asking for reactions to R.I.S.E. service and to the product was included with each packet of information that was sent out. Most clients returned the form, which, in general, showed a highly favorable response. Supplementing this effort were other evaluation procedures. These included on-site evaluations by outside experts as well as mandated evaluations by State evaluation groups. In addition, a random sampling of clients using typical interview techniques was conducted by the R.I.S.E. staff.

N.I.A., supported by private foundation funds, small fees from users, and Federal monies, is an educational service agency supplying research information to users drawn from throughout the county. Its information service employs optical coincidence scanning to serve the increasing number of their clients.

Purpose

The specific purpose of the comparison is to obtain some preliminary data to determine the difference in relevance of information received, as judged by the user. "Relevance" was defined as the relationship between a document and the user's information requirement.[2] While relevance judgments are complex processes,[3] it was decided to focus on this criterion because a demand literature search is the basic premise on which R.I.S.E. continues to serve its clients. Thus, "relevance" has more importance to R.I.S.E. than the issues of "speed," "cost," or "accuracy," which are considered secondary for purposes of this study.

A true experimental study, employing inferential statistical approaches, was not attempted.

Methodology

In a regular search of the E.R.I.C. collection, it is customary for R.I.S.E. to send to the client abstracts along with either excerpts of pertinent sections of a report or the complete report itself if not too lengthy.

It was decided, for purposes of the study, that only abstracts would be forwarded to the client by R.I.S.E. Since the N.I.A. sends only the accession number for any E.R.I.C. abstract, and not the abstract itself, to the requestor, a decision by R.I.S.E. to make a selection of excerpts from printed articles would have invalidated the study. The alternative of sending the complete printout for each item was ruled out on a financial and practical basis, since many E.R.I.C. documents run to several hundred pages.

It is known that there is a tendency to judge items more relevant if less is known about the subject and the person does not have sufficient evidence to decide between relevance and non-relevance.[4] Hence, it was known there would be instances where on the basis of an abstract alone, a client might make one decision regarding relevancy, but then change this decision after reading the complete report. Despite this, it was decided that providing only abstracts - while not an ideal method - was the most feasible course in a pilot study and further, that if a client erred because the abstract did not provide sufficient information, he would err in favor of both R.I.S.E. and N.I.A.

Four possible responses were decided upon: "relevant," "not relevant," "don't know but possibly relevant," or "don't know but possibly not relevant." The last two were selected because it was probable there would be a good number of items about which the client could not make a clear decision between "relevant" and "not relevant." It was deemed inappropriate to include only one additional possibility - "don't know" - since it is logical to expect certain instances where a client cannot decide between "relevant" and "don't Know" or between "not relevant" and "don't know." In such cases the client might check "relevant" instead of "don't know," not on the basis of the content of the abstract, but on the basis of past experiences with tests or evaluations, or of speculations about how the evaluation might be set up. It was believed that two choices within the "don't know" category might help solve these problems.

The possibility of providing a fifth choice - "don't know" - in addition to "don't know but possibly relevant" and "don't know but possibly not relevant" was considered.[5] It was ultimately decided that this was unnecessary since if a client could not make any decision about an abstract, it probably should be listed under "don't know but possibly not relevant," that is, both have the same meaning. Further, having a fifth choice would create a false value for choices three and four: without the fifth possibility, the client would have no qualms about checking "don't know but possibly relevant." If there were the fifth choice, however, the client might assume that choice three had "stricter criteria" than intended.

Procedure

In general, it was decided for purposes of the study, to accept the next 20 literature search requests in the order received by R.I.S.E. Excluded were certain types of requests so specialized that the topics would not likely be found in E.R.I.C. - for instance, patterns revealed by Rorschach tests; or because they were requests of a high priority nature, which necessitated unusual and hence a typical processing and treatment; or because they were requests submitted by clients under special contract. In the last case, it was felt that these clients might give biased responses because of the fee charged them.

Following acceptance of a search topic for the study, a detailed statement of the request was sent to the N.I.A., including a list of descriptors arrived at between the negotiator and the client.[6] At the beginning of the study, both agencies agreed that this procedure was mutually satisfactory and that no additional information on the question need be sought.

R.I.S.E. literature searchers were informed of the client's request in the usual way; they could ask any questions of interpretation of the supervisor who mediated the request initially. The searchers were asked to complete the E.R.I.C. portion of the search first and then fill out a special worksheet on

which to list accession numbers and so on, plus one extra copy of each applicable abst. ict. Searchers were informed this was part of a special pilot study, but were told to perform the E.R.I.C. portion of the search in the normal way.

When the abstract accession numbers were received from the N.I.A., these numbers were copied onto R.I.S.E.'s special worksheet. The searcher was then asked to make one copy of each abstract listed.

An evaluation form[7] was prepared for the client; it listed in numerical order all abstracts retrieved by both information agencies. The numbers from each were interdigitated so that the client would have no way of knowing which abstracts came from which source. If both agencies had selected the same abstract, only one entry was made on the evaluation form. The client also received one copy of each of the abstracts listed and was asked to judge the relevance on a subjective basis, that is, how a particular abstract related to his specific question not how it might relate to his subject in general. The clients were also informed that this pilot study was part of a special project and would interfere in no way with their regular ongoing search, the results of which would be mailed as soon as his evaluation form was completed and returned to R.I.S.E.

Upon receipt of the evaluation form the data were transferred to a summary sheet as follows: abstracts retrieved from both sources were listed first, along with the subjective responses; next came a listing of abstracts retrieved only by N.I.A.; the third group was the R.I.S.E. abstracts. The summary of results was tabulated at the bottom of the worksheet.

The pilot study was conducted from April to July 1969 and involved 20 respondents.[8]

Results

Table 1 presents data regarding the differences in relevance, as reported by users, between E.R.I.C. abstracts selected by R.I.S.E. and those selected

by the N.I.A.

Insert Table 1 about here

An examination of the data reveals, first, that R.I.S.E. searchers located almost twice as many abstracts as did N.I.A.: 118 or 68% of the total number of 173 abstracts on which clients expressed judgments for R.I.S.E., and 55 or 32% for N.I.A. Of the 118 abstracts identified by R.I.S.E., 56 or 47% were judged by users to be "relevant," and 26 or 22% as "not relevant." By contrast, of the 55 abstracts located by N.I.A., the clients checked 23 or 37% as "relevant," and 17 or 27% as "not relevant."

In terms of absolute figures, the comparison is even more striking. The users judged 56 abstracts located by R.I.S.E. as being relevant to their needs and only 26 as being not relevant, whereas they judged 23 abstracts identified by N.I.A. as relevant and 17 as being not relevant. In addition, the 20 clients reported 34 abstracts identified by R.I.S.E. as being "possibly relevant," compared to only 11 abstracts similarly identified by N.I.A. Clearly, then, the data show that the users judged many more R.I.S.E.-identified than N.I.A.-identified abstracts as being relevant.

From the viewpoint of individual cases, of 15 clients (four clients reported no relevant abstracts from either agency, and a fifth reported one relevant abstract from each agency), 11 reported more relevant abstracts located by R.I.S.E., while only 4 reported more relevant abstracts located by N.I.A.

Discussion

This study is subject to the usual limitations of studies involving a small number of cases; the reader is cautioned about the dangers of generalizing beyond the attempted purpose and limited scope of the study. Also, even by taking the requests for information in the order received, thus mirroring the "typical situation," variables, such as geographical proximity to the R.I.S.E.

Table I

Difference in Judged Relevance of R.I.S.E. and N.I.A. Abstracts

Client	R I S E					N I A					Total RISE and NIA
	R	N	DR	DN	Total	R	N	DR	DN	Total	
A	0	0	0	0	0	0	0	0	0	0	0
B	2	0	1	0	3	1	3	0	0	4	7
C	10	0	2	0	12	1	5	0	0	6	18
D	3	0	2	0	5	2	0	2	0	4	9
E	10	10	5	0	25	0	0	0	1	1	26
F	8	4	3	0	15	1	1	1	0	3	18
G	1	0	0	0	1	0	0	0	0	0	1
H	1	0	1	0	2	2	0	0	1	3	5
I	3	1	1	1	6	0	0	0	0	0	6
J	2	1	0	0	3	0	1	2	0	3	6
K	0	2	0	0	2	0	6	0	0	6	8
L	7	1	7	0	15	0	0	1	0	1	16
M	6	0	0	0	6	11	0	5	0	16	22
N	0	0	7	0	7	3	0	0	0	3	10
O	0	0	0	0	0	0	0	0	2	2	2
P	0	0	1	0	1	0	0	0	0	0	1
Q	1	1	1	0	3	1	0	0	0	1	4
R	1	4	2	1	8	0	0	0	0	0	8
S	0	0	1	0	1	1	0	0	0	1	2
T	1	2	0	0	3	0	1	0	0	1	4
	56	26	34	2	118	23	17	11	4	55	173

R = relevant
N = not relevant
DR = don't know but possibly relevant
DN = don't know but possibly not relevant

Note: Client "A" received abstracts, but he judged none to capable of being placed in any of the four response categories.

center, were not completely controlled for. On the other hand, interdigitating the abstracts and providing the same set of descriptors for both agencies did provide some measure of control.

Additional research is needed to confirm or deny these preliminary findings. For instance, a more statistically sophicated experimental study might be used as a follow-up of the limited pilot study. The response format could be designed so as to admit an examination by inferential statistical methods. Each response option could be weighed and a t test applied. Such refinements may yield greater precision. In the meantime, on the basis of the limited data collected and analyzed descriptively, R.I.S.E.- located (manually selected) abstracts were judged by users to be more relevant than N.I.A. (machine selected) abstracts.

Footnotes

[1]The fictitious name "National Information Agency" was chosen for purposes of the study.

[2]Donald King, "Design and Evaluation of Information Systems." In Annual Review of Information Science and Technology, Volume III, Encyclopedia Britanica, Inc., 1968, p. 70.

[3]Carlos Cuadra and Robert Katter, Special Libraries, "Implications of Relevance Research for Library Operations and Training" (Volume LIX, September 1968), p. 504.

[4]Ibid., p. 506.

[5]There was a point of view expressed in the early planning stages which argued for a fifth "don't know" choice, which would accomodate the client who definitely and honestly did not know. The provision for this fifth choice, it was held, would provide greater leeway on the continuem of a closure type response format. This argument was rejected in an initial pilot study for the reasons cited in the text. Perhaps a second research study should be made, employing a fifth choice.

[6]Robert S. Taylor, College and Research Libraries "Question-Negotiation and Information Seeking in Libraries" (May 1968). Mediating of all requests was done by a single full-time professional who was fully aware of the communication gaps" in the negotiation process. The average mediation, received over the telephone, took at least 20 minutes.

[7]See Appendix A.

[8]Of the 20 clients, 6 were teachers, 6 were directors, 3 were principals, 1 was a research intern, 1 was an elementary school supervisor, 1 was an assistant superintendent, 1 was a chief school administrator, and 1 was a college professor. There were 10 males and 10 females. In view of the small number of subjects, it was deemed not appropriate to analyze the data by staff position and/or sex.

APPENDIX A

R.I.S.E. SPECIAL EVALUATION _ ERIC MATERIALS

Name _____ _____ DATE

PROJECT TITLE _____

INSTRUCTIONS: The attached abstracts were selected by two different retrieval systems. The purpose of this evaluation is to provide a basis for judging the value of each system, as part of RISE's effort to improve its literature search services.

Please review each abstract and indicate your evaluation of its relevance, based on pertinence to your specific problem or question. Since these materials were prepared especially for your request, you are the only one who can judge their relevance.

These materials do not represent a completed literature search. Your project is in preparation and will be sent to you upon receipt of this form.

Check only one category for each abstract:

Abstract Number	Relevant	Not Relevant	Don't Know, But possibly Relevant	Don't Know, But possibly Not Relevant
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____
ED_____	____	____	____	____

omments, if any:

hank you for participating in this evaluation. We hope your efforts will be rewarded y improved RISE service in the future. Please return this form in the enclosed envelope o:

Supervisor of Dissemination
Research and Information Services for Education
443 South Gulph Road
King of Prussia, Pennsylvania 19406

Abstract

Differences in judged relevance of E.R.I.C. abstracts identified by two different information systems were investigated. A national semi-private educational service agency which utilizes mechanical searching of the research literature was one information system, while the other was a regional, Federally-funded agency employing manual searching. 20 clients reported they found substantially more E.R.I.C. abstracts identified by a manual method to be relevant to their needs than abstracts located by a mechanical searching system using the same set of E.R.I.C. descriptors.